

文章编号: 1001-0920(2013)10-1497-05

基于 K 均值聚类的二进制传感器网络多目标定位方法

黄月, 吴成东, 张云洲, 程龙, 孙尧

(东北大学信息科学与工程学院, 沈阳 110819)

摘要: 针对存在错误报警的二进制传感器网络, 提出基于 K 均值聚类的二进制传感器网络多目标定位方法. 在目标和节点间距离信息未知的条件下, 提出基于 K 均值聚类的改进加正减负算法(KMC-ISNAP)对目标位置进行估计, 引入影响因子降低分类过程中模糊节点对多目标定位误差的影响. 仿真实验表明, K 均值聚类方法在多个目标随机分布情况下能够对报警节点进行准确分类, 与质心估计算法和加正减负算法相比, KMC-ISNAP 多目标定位方法具有较高的定位精度和较好的容错性.

关键词: 无线传感器网络; 二进制传感器; K 均值聚类; 多目标定位

中图分类号: TP393

文献标志码: A

Multi-objective localization method based on K -means clustering in binary sensor networks

HUANG Yue, WU Cheng-dong, ZHANG Yun-zhou, CHENG Long, SUN Yao

(College of Information Science and Engineering, Northeastern University, Shenyang 110819, China. Correspondent: HUANG Yue, E-mail: yue.85@126.com)

Abstract: A multi-objective localization algorithm based on K -means clustering is proposed in binary wireless sensor networks with false alarm. The K -means clustering-improved subtract on negative add on positive(KMC-ISNAP) algorithm is applied to localize the multiple objectives where the distance between nodes and objectives is unknown, and influencing factors are used to reduce the influence of fuzzy nodes on localization errors. The simulation results show that the K -means clustering method is able to divide the alarmed sensors into parts accurately when multiple objectives are randomly distributed, and the proposed KMC-ISNAP has higher estimation accuracy and better fault tolerance than centroid estimator(CE) algorithm and subtract on negative add on positive(SANP) algorithm.

Key words: wireless sensor networks; binary sensor; K -means clustering; multi-objective localization

0 引言

无线传感器网络(WSN)由部署在监测区域内的大量传感器节点组成,其目的是协作地感知、采集、处理和传输感知对象的监测信息,目前已广泛应用于环境监测、医疗护理、智能家居、工业生产控制和军事侦查等领域.目标定位作为无线传感器网络的一项关键研究技术,是众多应用的基础^[1-3].

二进制传感器网络是一类特殊的WSN,该网络中的传感器节点只能提供目标存在或不存在于其探测范围内的单比特信息^[4-5].由于传感器节点提供的信息非常有限,二进制传感器网络目标定位与其他类型无线传感器网络定位算法有很大不同,是一个具有挑战性的问题^[6].文献[7]提出使用SNAP方法进行二

进制传感器网络单目标定位,具有较高的定位精度和较小的计算量.文献[8]提出了质心估计(CE)定位算法,在过滤掉瞬时故障引起的错误数据后计算所有报警节点坐标的平均值,估计目标位置,但当故障节点距离目标源较远时,该方法的误差较大.文献[9]提出使用极大似然估计方法计算二进制传感器网络中单气体源位置,具有较高的定位精度,但计算复杂度较高.文献[10]提出使用极大似然方法对多声源进行定位,该方法根据节点的超声能量观测模型建立基于测量距离的多源定位模型,但需要每个传感器节点传送的数据量较大,缩短了节点的生存时间.文献[11]提出基于信任指标的故障容忍多事件定位算法,通过为每个节点分配一个0~1之间的信任指数降低误警节

收稿日期: 2012-10-09; 修回日期: 2013-03-04.

基金项目: 国家自然科学基金项目(61273078); 国家重点实验室基金项目(RLO200913).

作者简介: 黄月(1985-),女,博士生,从事无线传感器网络定位与部署的研究; 吴成东(1960-),男,教授,博士生导师,从事无线传感器网络、图像处理等研究.

点对定位精度的影响,能够取得较好的定位效果,但信任指数的分配和更新依赖于采样观测值。

WSN单目标定位算法相对比较成熟,但多目标定位算法是研究中的难点,而在节点和目标间距离信息未知的二进制传感器网络中,多目标定位更加难以实现。WSN在实际应用中可能由于环境影响、电池耗尽和节点故障等因素导致节点错误报警(包括漏报和误报),本文针对存在错误报警的二进制传感器网络提出了基于 K 均值聚类的多目标定位方法,在目标和节点间距离信息未知的条件下,通过 K 均值聚类方法对报警节点进行分类,并引入影响因子降低节点对定位精度的影响。仿真结果表明,所提出算法具有较高的定位精度和较好的容错性,适用于存在错误报警的二进制传感器网络多目标定位。

1 系统模型

本文的研究基于以下假设:

1) 传感器节点随机分布在监测区域内,且已知每个节点位置信息。

2) 所有传感器节点具有相同的探测半径,且探测半径大小已知。

3) 监测区域内存在多个位置随机分布的待定位目标,目标个数已知。

4) 目标源持续不断地发出超声波信号,且向各个方向传播一致,传播延迟可以忽略。

5) 存在两种故障节点,一种节点在目标影响区域内并未发生报警,保持沉默状态;另一种节点在目标影响区域外发生错误报警,并向基站发送信号1。

假定目标源发出的信号衰减与传感器节点到目标距离的 $\alpha(\alpha \in R^+)$ 次幂成反比,位于 (x_n, y_n) 的传感器节点 n 的观测值模型为

$$Z_n = \min \left\{ A_{\max}, \gamma \sum_{k=1}^K S_{n,k} + \omega_n \right\},$$

$$n = 1, 2, \dots, N; \quad (1)$$

$$S_{n,k} = w_k / R_{n,k}^\alpha. \quad (2)$$

其中: A_{\max} 为传感器的最大测量范围; γ 为传感器增益; K 为待定位的目标源个数; w_k 为第 k 个目标源的信号强度; ω_n 为系统噪声; $R_{n,k}$ 为节点 n 到目标 k 的距离,且有

$$R_{n,k} = \sqrt{(x_n - x_k)^2 + (y_n - y_k)^2}. \quad (3)$$

每个信标节点被预先设定了相同的阈值 T (阈值足够大),依此确定报警和非报警节点:若 $Z_n \geq T$,即观测值大于等于阈值的节点,则为报警节点;若 $Z_n < T$,即观测值小于阈值的节点,则为非报警节点。

2 基于 K 均值聚类的节点分类

K 均值聚类算法作为传统的动态聚类方法,广泛

应用于科学和工业等领域,具有简单、快速高效等诸多优点^[12]。本文采用 K 均值聚类算法对报警节点进行聚类,使之与相应目标源对应。输入量 K 为目标源个数, N 为报警节点个数,聚类后的每个 v_j 对应一个待定位目标。聚类模型如下: $X = \{X_1, X_2, \dots, X_N\}$ 为 N 个待分类样本的集合, X 的一个分划为 $V = \{v_j | j = 1, 2, \dots, K\}$,满足

$$\begin{cases} X = \bigcup_{j=1}^K v_j, v_j \neq \emptyset, j = 1, 2, \dots, K; \\ v_i \cap v_j = \phi, i, j = 1, 2, \dots, K, i \neq j. \end{cases} \quad (4)$$

其中 v_j 为分划 V 中的类。聚类的目的是找到 X 的一个分划 V^* ,使 V^* 的类内离散度和 $J_{V^*} = \sum_{j=1}^K \sum_{x \in v_j} d(x, c_j)$ 达到最小。其中: c_j 为第 j 个聚类中心; $d(x, c_j)$ 为聚类内的数据对象到聚类中心的欧氏距离; J_{V^*} 为聚类准则函数,表示所有样本到各自聚类中心的距离总和。在某次迭代中,当所有样本被正确分类时,聚类中心不会再改变,表明聚类准则函数已经收敛,算法结束。

基于 K 均值聚类算法的节点聚类伪代码如下。

Step 1: 网络初始化。

Step 2: 从 N 个数据对象中任意选择 K 个对象作为初始聚类中心。

Step 3: 计算节点到每个聚类中心的欧氏距离 D ,按最小距离准则分配给 K 个聚类的某一个,

for $i = 1 : N$

$$n_j \in v_j; \text{ if } d(n_j, c_j) = \min D$$

end

Step 4: 计算 K 个聚类新的中心,

for $j = 1 : K$

$$x_{c_j} = \frac{1}{q} \sum_{i=1}^q x_{\text{alarm}}(i), y_{c_j} = \frac{1}{q} \sum_{i=1}^q y_{\text{alarm}}(i),$$

end

Step 5: 返回 Step 3, 将报警节点重新聚类,重复迭代运算。

Step 6: 若聚类 v_j 元素不再变化,则算法收敛,计算结束。

3 多目标定位算法

3.1 CE 算法

CE算法中,对聚类后报警节点的坐标求平均值,该值为所求目标源的位置坐标,即

$$x_{c_j} = \sum_{i=1}^{v_j} x_{\text{alarm}}(i), y_{c_j} = \sum_{i=1}^{v_j} y_{\text{alarm}}(i),$$

$$j = 1, 2, \dots, K.$$

该算法估计的精确度与信标节点密度和分布有关,且易受误警节点影响.

3.2 SNAP 算法

加正减负算法 (SNAP) 是一种极大似然估计方法, 通过节点汇报的数据构造似然函数. 指示函数如下:

$$I_n = \begin{cases} 0, & Z_n < T; \\ 1, & Z_n \geq T. \end{cases} \quad (5)$$

其中 I_n 为节点汇报的数据, $I_n \subseteq I = \{I_n : R_{n,k} \leq R_c\}$, $R_{n,k}$ 为节点 n 到目标源 k 的距离. 似然函数为

$$\log p(I_n|\theta) = \sum_{n=1}^N I_n \times \lg \left[Q\left(\frac{T - S_n(\theta)}{\sigma_\omega}\right) \right] + (1 - I_n) \times \lg \left[1 - Q\left(\frac{T - S_n(\theta)}{\sigma_\omega}\right) \right]. \quad (6)$$

文献 [8] 给出如下概率值:

$$P_r\{I_n = 1|\theta\} = Q\left(\frac{T - S_n(\theta)}{\sigma_\omega}\right) = 0.99, \quad (7)$$

$$P_r\{I_n = 0|\theta\} = Q\left(\frac{T - S_n(\theta)}{\sigma_\omega}\right) = 0.01, \quad (8)$$

其中 θ 为目标源位置. 根据修正的似然函数 $p'(I_n|\theta) = 10^{2n} p(I_n|\theta)$, 对其取对数得到

$$\begin{aligned} \log p'(I_n|\theta) &= \sum_{n=1}^N I_n \times \log(9.9) + (1 - I_n) \times \log(0.1) \approx \\ &= \sum_{n=1}^N I_n \times (+1) + (1 - I_n) \times (-1), \end{aligned} \quad (9)$$

则 SNAP 估计为

$$\hat{\theta}_{\text{snap}} = \max \log p'(I_n|\theta). \quad (10)$$

3.3 KMC-ISNAP 算法

SNAP 算法利用传感器节点汇报的二进制观测数据对目标进行定位, 其基本思想是对监测区域进行网格划分后构造似然矩阵, 根据节点汇报的二进制观测值对以该节点为中心的一个固定区域对应的矩阵元素进行加 1 或减 1, 最后, 矩阵中数值最大的元素对应的位置即为目标源位置. 该算法可以有效减小误警节点的影响, 在单目标定位时有较好的定位效果.

本文针对 SNAP 算法作出改进, 提出了适合多目标定位的基于 K 均值聚类的改进加正减负算法 (KMC-ISNAP), 描述如下.

Step 1: 根据节点汇报的二进制数据采用 KMC 算法对报警节点进行聚类.

Step 2: 将 WSN 监控区域划分成 $M \times M$ 的网格, 网格分辨率不同, 方形区域的大小不同, 即覆盖监控区域单元格的数目不同.

Step 3: 根据划分的网格单元构造一个矩阵 L , 每个单元格对应一个矩阵元素, 即

$$L = \begin{bmatrix} L_{11} & L_{12} & \cdots & L_{1m} \\ L_{21} & L_{22} & \cdots & L_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ L_{m1} & L_{m2} & \cdots & L_{mm} \end{bmatrix}, \quad (11)$$

其中 L_{ij} 为在 (i, j) 处的网格单元值.

Step 4: 利用聚类 v_1 中节点对矩阵元素更新, 有

$$L_{v_1}^{h+1}(i, j) = \begin{cases} L_{v_1}^h(i, j) + 1 \times e(i, j), & d_{\text{alarm}}(i, j) \leq R_c; \\ L_{v_1}^h(i, j) - 1 \times e(i, j), & d_{\text{unalarm}}(i, j) \leq R_c; \\ 0, & \text{otherwise.} \end{cases} \quad (12)$$

$$e(i, j) = \begin{cases} 0.8, & \Delta d(i, j) \leq U; \\ 1, & \Delta d(i, j) > U. \end{cases} \quad (13)$$

其中: $d_{\text{alarm}}(i, j)$ 为网格中心与 v_1 中报警节点的欧氏距离; $d_{\text{unalarm}}(i, j)$ 为网格中心与 v_1 中未报警节点的欧氏距离; R_c 为传感器节点的探测半径; $e(i, j)$ 为与任意两个聚类中心距离差小于 U 的报警节点对矩阵元素的影响因子, 取 $U = 0.2R_c$; Δd 为报警节点与聚类中心的距离差, 将距离任意两个聚类中心距离差较小的报警节点定义为模糊节点, 模糊节点在分类过程中虽然划分为最近的一类, 但也有可能是感知到其他目标源而发生报警. KMC-ISNAP 算法中的影响因子可以降低模糊节点对多目标定位的不利影响, 有利于提高定位精度.

Step 5: 计算矩阵元素最大值, $\max L(i, j)$ 对应的单元格网格中心为估计的目标 1 位置, 若有多个矩阵元素值相等, 则这些矩阵元素对应的单元格质心为估计目标 1 位置.

Step 6: 对于其他聚类, 重复 Step 4 和 Step 5, 估计下一个目标的位置.

4 仿真分析

本文提出的基于 K 均值聚类的多目标定位算法在 Matlab 仿真环境中实现, 实验参数如表 1 所示.

表 1 实验相关参数值

参数	取值
信标节点个数	100
节点探测半径	25
信号源强度	3000
噪声模型	$\omega_n \sim N(0, 1)$
传感器增益	2
报警阈值	5

对于随机分布在 WSN 监测区域的多个目标源, 假定目标随机分布的位置为 (x_k, y_k) , 通过定位算法求出的目标位置为 (\hat{x}_k, \hat{y}_k) . 本文利用多次实验的均方根误差作为定位算法的性能评价标准, 表达式如

下:

$$\text{RMS}_{\text{error}} = \frac{1}{K} \sqrt{\frac{\sum_{i=1}^B \sum_{k=1}^K (x_k - \hat{x}_k)^2 + (y_k - \hat{y}_k)^2}{B}}, \quad (14)$$

其中 B 为实验次数. 本文取 $B = 1000$, 每次实验中传感器节点分布固定, 目标源随机分布.

当网络中存在两个待定位的目标源时, 节点接收到目标源发出的信号后报警并向基站发送信号 1, 其中包括探测到目标 1、探测到目标 2、探测到两个目标和未探测到目标的误警. 根据节点观测信息采用 KMC 算法对报警节点进行聚类后再进行定位估计. 多个目标源时的聚类成功率如图 1 所示. 由图 1 可见, 监控区域内存在探测到目标而并未发生报警的漏报节点和未探测到目标而发生报警的误报节点, 基于 KMC 的聚类方法可以将报警节点有效分类, 使之与相应目标源对应.

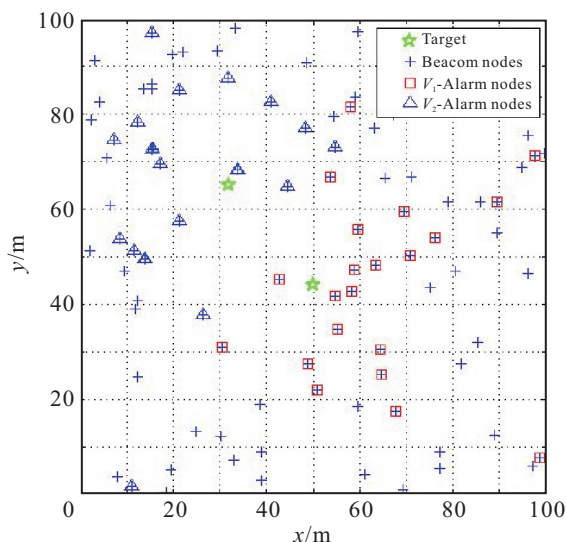


图 1 基于 K 均值聚类的报警节点分类效果

表 2 为多个目标源时的聚类成功率. 由表 2 中数据可知, 当网络中存在的待定位目标源个数增加时, 基于 KMC 算法的聚类成功率略有降低, 但聚类成功率均在 90% 以上, 对定位结果基本没有影响.

表 2 多个目标源时的聚类成功率

目标源数目/个	2	3	4	5	6
聚类成功率/%	93.99	92.79	91.46	90.45	90.02

图 2 为目标源个数不同时的定位误差. 由图 2 可见, 随着目标源个数的增加, 定位误差逐渐增大, 因为当目标源个数增加时, 每个聚类中的报警节点数目变少, 所以导致定位误差的增大. 当网络中的信标节点数目增加时, 每个聚类中的报警节点数目也会相应增多, 定位误差减小. 当网络中待定位的目标源增多时,

定位问题将变得更加复杂, 本文算法采用先聚类再定位的方式完成多目标定位, 计算复杂度不受目标源数目的影响.

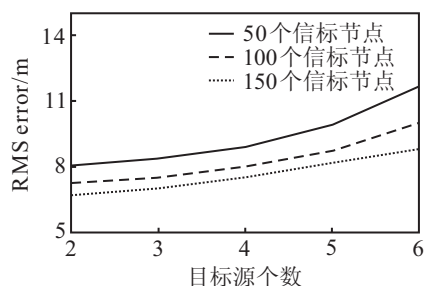


图 2 不同目标源时的定位误差

当目标源个数相同时, 分别采用 CE 算法、SNAP 算法和 KMC-ISNAP 算法进行定位估计. 图 3 为误警节点比例对定位算法的影响. 由图 3 可见, 随着误警节点比例的增加, KMC-ISNAP 和 SNAP 算法定位误差变化较小, CE 算法的定位误差最大. SNAP 算法未考虑离两个信号源都较近的模糊节点对定位效果的不利影响; KMC-ISNAP 算法引入影响因子降低了模糊节点对定位的影响, 具有较高的定位精度和较好的容错性. 随着误警节点的增多, CE 算法的定位误差发生较大变化, 该算法在理想情况下具有较好的定位效果, 但受误警节点影响也较大.

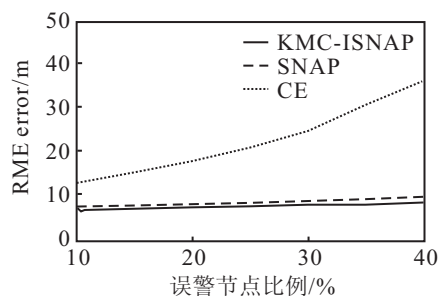


图 3 误警节点比例对定位算法的影响

图 4 为信标节点数目对定位算法的影响. 由图 4 可见, 3 种算法的定位误差随着信标节点数量的增加而减小, 这是由于节点密度增加后, 减小了误警节点对定位误差的不利影响. 信标节点数目相同时, CE 算法的定位误差最大, KMC-ISNAP 算法的定位误差最小.

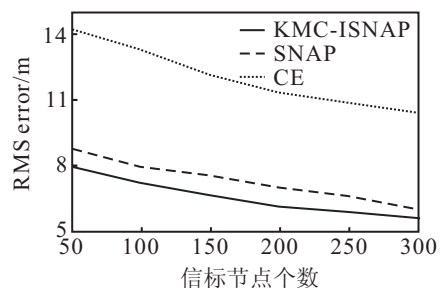


图 4 信标节点数目对定位算法的影响

图 5 为单元格长度对定位算法的影响. 由图 5 可见, SNAP 和 KMC-ISNAP 算法的定位误差随着单元格长度变大而增大, CE 算法定位误差与单元格长度无关, 相同单元格长度下, KMC-ISNAP 算法的定位误差最小. 单元格长度增大, 则单元格个数变少, 矩阵维数减小, 计算量降低, 实际应用中可以通过权衡定位精度和计算复杂度来设置单元格长度.

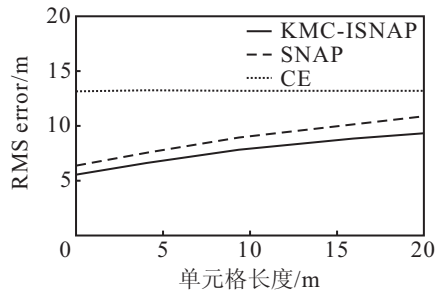


图 5 单元格长度对定位算法的影响

5 结 论

本文针对存在错误报警的二进制传感器网络, 提出了基于 K 均值聚类的二进制传感器网络多目标定位算法. 根据二进制节点的观测信息, 采用 K 均值聚类方法对报警节点进行分类, 使报警节点与相应的目标源对应. 在目标和节点间距离信息未知的条件下, 提出了 KMC-ISNAP 算法对目标源进行定位估计, 引入了影响因子以降低分类过程中的模糊节点对定位精度的影响. 通过仿真实验表明了, 在多个目标源随机分布的情况下, 本文提出的基于 K 均值聚类的二进制传感器网络多目标定位算法在不同信标节点个数、不同误警节点比例和不同单元格长度时, 与 CE 算法和 SNAP 算法相比较, 具有较高的定位精度和较好的容错性.

参考文献(References)

- [1] Chiu Wei-yu, Chen Bor-sen, Yang Chang-yi. Robust relative location estimation in wireless sensor networks with inexact position problems[J]. IEEE Trans on Mobile Computing, 2012, 11(6): 935-946.
- [2] Gasparri A, Pascucci F. An interlaced extended information filter for self-localization in sensor networks[J]. IEEE Trans on Mobile Computing, 2010, 9(10): 1491-1504.
- [3] Gharehshiran O N, Krishnamurthy V. Coalition formation for bearings-only localization in sensor networks — A cooperative game approach[J]. IEEE Trans on Signal Processing, 2010, 58(8): 4322-4338.
- [4] 孙晓艳, 李建东, 陈彦辉, 等. 二进制传感器网络加权目标跟踪算法研究[J]. 电子与信息学报, 2010, 32(9): 2052-2057.
(Sun X Y, Li J D, Chen Y H, et al. The study on weighted target tracking algorithm for binary sensor networks[J]. J of Electronics and Information Technology, 2010, 32(9): 2052-2057.)
- [5] Razi A. Binary source estimation using a two-tiered wireless sensor network[J]. Communications Letters, 2011, 15(4): 449-451.
- [6] 周德超, 吴晓平. 二进制无线传感器网络目标定位的解析算法[J]. 武汉理工大学学报, 2008, 30(8): 78-81.
(Zhou D C, Wu X P. Analytic arithmetic for target locating with binary wireless sensor networks[J]. J of Wuhan University of Technology, 2008, 30(8): 78-81.)
- [7] Ding M. Fault-tolerant target localization in sensor networks[J]. Eurasip J on Wireless Communications and Networking, 2007, 1: 1-9.
- [8] Michaelides M P, Panayiotou C G. SNAP: Fault tolerant event location estimation in sensor networks using binary data[J]. IEEE Trans on Computers, 2009, 58(9): 1185-1197.
- [9] Vijayakumaran S, Levinbooy Y, Wong T F. Maximum likelihood localization of a diffusive point source using binary observations[J]. IEEE Trans on Signal Processing, 2007, 55(2): 665-676.
- [10] Sheng X H, Hu Y H. Maximum likelihood multiple-source localization using acoustic energy measurements with wireless sensor networks[J]. IEEE Trans on Signal Processing, 2005, 53(1): 44-53.
- [11] Xu Xiang-hua, Gao Xue-yong, Wan Jian, et al. Trust index based fault tolerant multiple event localization algorithm for WSNs[J]. Sensors, 2011, 11(7): 6555-6574.
- [12] Shi Yu, Tranchevent L C, Xinhai Liu, et al. Optimized data fusion for kernel k -means clustering[J]. IEEE Trans on Date of Publication, 2012, 34(5): 1031-1039.
- [16] Qu Z H. Cooperative control of dynamical systems: Applications to autonomous vehicles[M]. New York: Springer-Verlag, 2009: 167-171.
- [17] Yu S H, Yu X H, Shirinzadeh B, et al. Continuous finite-time control for robotic manipulators with terminal sliding mode[J]. Automatica, 2005, 41(11): 1957-1964.

(上接第1496页)